A protein nucleic search using tfasta_plus_p2n model

Run on:

SeqCore version 6.1.4-p5-4578
Copyright (c) 1993 - 2003 CoGene Ltd.

May 21, 2003, 00:27:44 : Search time 1713 Seconds
(without alignments)
1512.713 Million Cel. updates/sec

Title: US-10-017-754-809
Perfect score: 887
Sequence: 1 MRCHARGFSCLVTACREES..................IIPKEEHCKMPEAGEDPGV 160

Scoring table: BLOSUM62
             Xgapop 10.0 , Xgapext   0.5
             Ygapop 10.0 , Ygapext   0.5
             Fgapop 6.0 , Fgapext   7.0
             Delop 7.0 , Delext   7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 42258132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
             Maximum Match 100%
             Listing first 45 summaries

Command line parameters:
MODEL=tframe+_p2n.model -Dev-xlh
H=/cg2.1/USPTO_spool/US100:1754/runat_1405_2003_160855_1031/app_query.fasta : 327
DB-EST -QFMT-fastp -SUFFIX-rst -MINMATCH-1 -LOOPCL-1 -LOOPEXT-1
UNITS-bits -START-1 -END--1 -MATRIX-blosum62 -TRANS-hu:aa63.cdi -LIST-45
LOCALIGN-200 -THR_SCORE-pct -THR_MAX-100 -THR_N-0 -ALIGN-15 -MODE-LOCAL
OUTFMT=pto -NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-200000000
USER=US10:01754_3CGN_1.1_1906_3runat_1405_2003_160855_1031 -NCPU-6 -ICPU-3
IC_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES-9 -WAIT -CNGLOG -DEV -TIMEOUT-120
ARN_TIMEOUT=30 -THREADS-1 -XGAPOP-10 -XGAPEXT-5
GAPOP=10 -YGAPEXT-0.5 -DELOP-0 -DELEXT-7

Database :   EST.*
             1: em_estba:*
             2: em_esthum:*
             3: em_estin:*
             4: em_estro:*
             5: em_estmu:*
             6: em_estp.:*
             7: em_estro:*
             8: em_brc.:*
             9: cm_est1:*
            10: qb_est2:*
            11: gb_htc:*
            12: qb_est3:*
            13: qb_est4:*
            14: gb_est5:*
            15: em_estrum:*
            16: em_estom:*
            17: qb_ass:*
            18: em_gss_hum:*
            19: em_gss_lov:*
            20: em_estmu:*
            21: em_gss_pln:*
            22: em_gss_vrt:*
            23: em_gss_fun:*
            24: em_gss_ram:*
            25: em_gss_mus:*
            26: em_gss_other:*
            27: em_gss_pro:*
            28: em_gss_rod:*

SUMMARIES

Result No.	Score	Query Match Length	Sp Db	Accession	Description

RESULT:
BB975341
LOCUS   BB975341
DEFINITION BB975341 NIH_MGC_69 Homo sapiens cDNA clone IMAGE
            mRNA sequence.
ACCESSION   BB975341
VERSION   BB975341.1
KEYWORDS   EST
SOURCE   human
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 785)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)

```
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: the I.M.A.G.E. Consortium
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: IIAM676  row e  column 13
            High quality sequence stop: 724
FEATURES             Location/Qualifiers
     source          1..785
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3949444"
                     /clone_lib="NIH_MGC_69"
                     /tissue_type="large cell carcinoma, undifferentiated"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned undirectionally; Primer: oligo dT;
                     Average insert size 1.1 kb. Library constructed by Life
                     Technologies."

S-10-017-754-809 (1-160) x BE876343 (1-785)

Alignment Scores:
Pred. No.:
Score.:                  2.016e-69          Length:         785
Percent Similarity:      845.20             Matches:        157
Percent Local Similarity:99.74%            Conservative:   0
Query Match:             95.26%            Mismatches:     2
                         52                Indels:         2
                                           Gaps:           2

ORIGIN
         175 c    234 g    133 t

  1 MetArgCysHisAlaHisGlyProSerCysLeuValThrAlaLeuThrAlaHisGly 20
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
163 ATGCGGTGCCACGGCCCATGGACCCTCTTGTCACGTCACGGCCATACGAAGGGGAGA 222

 21 GlyProArgSerGlyGlyAlaGlyIleAlaLysLeuSerCysCysSerGlyTyrTyrProSerPro 40
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
223 GGGCCGAGGAGTGGAGGGCGTCAGGTGAAGCCTGGGTGCTGTTCGGGGTATCCGAGCCC 282

 42 ArgSerThrIleProAsnProGluArgThrProGlnLysArgGlyProGlyCysGly 60
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
283 ACAGAGCACCCTGCAACCCCGACCAGAAGCAICTCGGCTCCCCAGAGGACCAGGAAGGGGA 342

 61 ArgHisSerGlnArgThrGlnThrGlnAsnHisIleAlaSerProArgSerProValMet 80
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
343 CGGCATGAGCCGCACACACACAAACACAGAACCACAGCCAGTCCAGGAGCCCAGTAA 412

 81 GlnSerProLysHisGlnSerGlnIleAspGluArgSerAlaGlyIleLeuSerGlySerArg 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
413 GACAGCCCCAAAAACAGAGCAGCAGAAAGCCAGATCGAGGTCCCACACGAAGGCATGCGAGA 462

 101 GluLysLysIleCysArgIleProIleArgAsnHisIleAlaSerProArgSerProVal 120
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
463 CAGAAGAAGATCAGATAGCCTCATACCCAGTGGCCACAATGGAAGTCAGTCAAA 522

 121 SerArgCysGlnGluGlnThrProArgTyrIleAsnLeuAspLeuGlySerArgGlyValValGlySerVal 140
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
523 SAGCCIGAICACIGTCAAACACCGGCGATAAAATCTGAATCGCAGAAACGTGCAAGCAA 582

 140 TyrSerIleLeuProGlySerGluHisSerCysHisMetProSerLeuAlaGlyLysIleLeuSer 159
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
583 AGAATAATACCTAAGAGGCAAACACTGAAAATGCCAGAAGCCAGTGCAAGAGCAA 635

LOCUS       BE881023               489 bp    mRNA        linear   EST 20-OCT-2000
DEFINITION  601490780F1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3949227 5'
```

```
LOCUS       BE881023               489 bp    mRNA        linear   EST 20-OCT-2000
ACCESSION   BE881023
VERSION     BE881023.1
KEYWORDS    EST
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homo.
REFERENCE   1 (bases 1 to 489)
  AUTHORS   NIH-MGC Project.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: the I.M.A.G.E. Consortium
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: IIAM680  row d  column 20
            High quality sequence stop: 489
FEATURES             Location/Qualifiers
     source          1..489
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3949227"
                     /clone_lib="NIH_MGC_69"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned undirectionally; Primer: oligo dT;
                     Average insert size 1.1 kb. Library constructed by Life
                     Technologies."

BASE COUNT
ORIGIN
         163 a    111 g    85 t

S-10-017-754-809 (1-160) x BE881023 (1-489)

Alignment Scores:
Pred. No.:
Score:                   1.91e-60           Length:         489
Percent Similarity:      747.00             Matches:        135
Percent Local Similarity:100.00%           Conservative:   0
Query Match:             84.22%            Mismatches:     0
                         12                 Indels:         0
                                            Gaps:           0

  1 MetArgCysHisAlaHisGlyProSerCysLeuValThrAlaLeuThrAla 17
    ||||||||||||||||||||||||||||||||||||||||||||||||
 26 GGCAGGCCGCATGGACCCTCTTGTCACGTCACGGCCATACGAAGGGGAGA 81

 18 ProAspArgGlnAsnHisArgArgIleGluTyrLeuSerSerProArgSerPro 38
    ||||||||||||||||||||||||||||||||||||||||||||||||
  1 CCCGACAGAAGCAGAAACCACAGGAATATCTGAGCAGTCCAAGGAGCCCAGTA 81

 39 ValArgLysIleLeuGlySerProThrSerGlnAsnHisIleAlaSerProArgSer 57
    ||||||||||||||||||||||||||||||||||||||||||||||||
 82 GTCAGGAAGATCCTAGGAAGTCCCACCAGTCAGAATCACATTGCAAGTCCAAGAAGT 141

 58 ProValAsnHisArgArgIleLeuGlySerProThrSerGln 66
    ||||||||||||||||||||||||||||||||||||||||||||
142 CCGGTTAACCATCGGCGGATCTTGGGCAGTCCGACAAGTCAGAA 242

 67 LysIleThrValArgSerArgGlnGlyValValGlyArgValMetAlaSerProArgSerPro 85
    ||||||||||||||||||||||||||||||||||||||||||
 66 AAGATAACAGTCAGATCCAGACAAGGAGTAGTAGGGAGAGTAATGGCTAGTCCGAGAA 125

 86 ValLysIleSerArgSerLeuSerProValLysLysIleThrValLysSerProValLeu 105
    ||||||||||||||||||||||||||||||||||||||||||
126 GTAAAGATAAGTCGATCTCTGAGCCCAGTGAAAAAGATAACAGTGAAGTCA 285

106 LeuArgSerArgGlnGlyValValGlySerValLysLysIleLeuProGlySerGlu 122
    ||||||||||||||||||||||||||||||||||||||||||
242 ACTCCGAGAAGCCGACAAGGAGTAGTAGGGAGTGTAAAAAAGATTTTACCTGGCAGTGAA 307

123 HisSerCysHisMetProSerLeuAlaGlyLysIleLeuSerProGlyValVal 140
    ||||||||||||||||||||||||||||||||||||||||||
302 ACACTGCAAGCATGCCGAGCATTCGCTGGCAAGATCCTGAGTCCAGGTGTGGTAA 357
```

RESULT 3

362 CAACCACTG~AAAATGCCGAAAGAAGTGAAGAGTAA~CACAA~T~ 4 6

AA758618
LOCUS   AA758618/c
DEFINITION a158d07.s1 Soares_testis_NHT Homo sapiens cDNA clone ...
            mRNA sequence.
ACCESSION   AA758618
VERSION   AA758618.1   GI:2806481
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 407)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov

FEATURES   Location/Qualifiers
     source   1..407
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"

ORIGIN

LOCUS   AI631681   421 bp   mRNA   linear   EST 17-FEB-1999
DEFINITION wb08d02.x1 NCI_CGAP_Co6 Homo sapiens cDNA clone IMAGE:2305059 3',
            mRNA sequence.
ACCESSION   AI631681
VERSION   AI631681.1 GI:4638301
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 421)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Insert Length: 515   Std Error: 0.00
            Seq primer: -40P from Gibco.
FEATURES   Location/Qualifiers
     source   1..421
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2305059"
                     /clone_lib="NCI_CGAP_Co6"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site 1, Not I; Site 2, Eco R1; plasmid DNA
                     from the normalized library NCI_CGAP_Co4 was prepared and
                     ss circles were made in vitro. Po-Gwisa HAP purification.
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. the driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (clone ID
                     1257035-1255031, 1469564-1470483, and 1475592-1476499).
                     Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT   75 a   95 c   95 g   156 t
ORIGIN

Alignment Scores:
Prid. No.:   7.62e-99
Score: 516.00
Percent Similarity: 98.00%
Best Local Similarity: 98.00%
Query Match: 58.178%

Length: 421
Matches: 98
Conservative: 3
Mismatches: 2
Indels: 0
Gaps: 0

US-10-017-754-809 (1-160; × Ala-Ile-Ala; 1-421)

LOCUS   AI631681   421 bp   mRNA   linear   EST
DEFINITION U.S. NCI_CGAP_Co6 Homo sapiens cDNA clone IMAGE:2305059 3',
            mRNA sequence.
ACCESSION   AI631681
VERSION   AI631681.1 GI:4638301
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 421)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES   Location/Qualifiers
     source   1..428
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2305059"
                     /clone_lib="NCI_CGAP_Co6"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site 1, Not I; Site 2, Eco R1; plasmid DNA
                     from the normalized library NCI_CGAP_Co4 was prepared and
                     ss circles were made in vitro. Po-Gwisa HAP purification.
                     Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT   104 a   93 c   93 g   138 t
ORIGIN

Alignment Scores:
Prid. No.:   7.76e-99
Score: 516.00
Matches: 98

Percent Similarity: 98.094   Conservation: 2
Best Local Similarity: 98.094   Mismatches: 2
Query Match: 58.178   Indels: 0
%:   Gaps: 0

US-10-017-754-809 (2-160) x AA448244 (1-428)

Qy   61   ArgHisGluArgHisThrGlnThrGlnAsnHisThrArgAsnHisProSerValMet
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   145 AGACACGAGAGACACAACACAAAACACACAGAACATCACACGAATCACCCAAGCATG

Qy   91   SerSerProLysLysLysAsnGlnGlnIleuLysValIleu...GlnHisLeuProSerArg
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   329 GAGAGCCCCAAAAAGAAGAAACCAGCAGCTGAAGTCCGGATC

Qy   121 SerGlyIleSerGlnThrArgGlyIleuAsnLeuAspLeu...ValLysVal
           |||||||||||||||||||||||||||||||||||||
Db   209 AGCTGCATCAGTCAAAACACCGGGATAAAATCTGAAGTC

Qy   141 IleIleProLysGluGlnHisCysLysMetProGlu...
           ||||||||||||||||||||||||||||||
Db   149 ATAATACCTAAAGAGGAACACTGTAAAATG

RESULT 7
LOCUS   AA954712
DEFINITION   wq32c10.x1 NCI_CGAP_GC5 Homo sapiens cDNA clone IMAGE:2472979 3'
ACCESSION   AA954712
VERSION   AA954712.1
KEYWORDS   EST
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 438)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
             R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
             Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www.bio.llnl.gov/bbrp/image/image.html
             Insert Length: 536   Std Error: 0.00
             Seq primer: -40up from Gibco
             High quality sequence stop: 396
FEATURES   Location/Qualifiers
     source   1..438
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2472979"
                     /tissue_type="pooled germ cell tumors"
                     /clone_lib="NCI_CGAP_GC5"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site 1: Not I; Site 2: Eco RI; plasmid DNA
                     from the normalized library NCI_CGAP_GC4 was prepared and
                     ss circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (clone IDs
                     1257096-1258631, 1469064-1470983, and 1475592-1476743)."

BASE COUNT
ORIGIN

normalized to POT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci) in preparation, library
constructed by M. Brownstein (N:MH/NHGRI; National
Institutes of Health). Note: this is a NIH_MGC library."

LOCUS   BG772292
DEFINITION 6G2722580F1_NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4839426 5',
            mRNA sequence.
ACCESSION   BG772292
VERSION   BG772292.1 GI:14082945
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 487)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NIGMS), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Lucigen Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov/
            Plate: ILAM0274 row: P column: 19
            High quality sequence stop: 463
FEATURES   Location/Qualifiers
     source   1..487
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4839426"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"

/clone "IMAGE:3889095"
/clone_lib "NIH_MGC_65"
/tissue_type="large cell carcinoma; adifferentiated"
/lab_host "DH10B (phage-resistant)"
/note "Organ: lung. Vector: pCMV-SPORT6. Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Priming method:
Average insert size 1.1 kb."
Technologies."

BASE COUNT   382 a   192 c   204 g   97 t
ORIGIN

S-10-017-754-809 (1-160) x BG472234 (1-889)

Alignment Scores:
Score: 1.94e-38   Length: 889
Percent Similarity: 98.00%   Matches: 98
Best Local Similarity: 98.00%   Conserved: 98
Query Match: 58.27%   Mismatches: 2
                       12   Indels: 0
                                        Gaps: 0

62 ArgHisGlnArgHisThrGlnAsnHisThrAlaSerProArgSerProValMet...
54 AGACACGGGCGGCACACACNAAACACNAACACNAGATCCAGTCCCAGAGCCCCATANTC...

81 GluSerProLysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisSerGlyerArg
114 GAGAGCCCAAAAAGAAGAAACCAGNGCTGAAAGTGGGATCCNAGTCCAGGCAGCA...

101 GlnGlyLysLysLeuArgIleLeuLeuArgSerGlnGlyAlaIleTrpLysCysLys
174 TAGAGAGAAGANTCAGGATACNGCTGAGATCCCAGTCCGACATGAAGCTGANTCGAAG

121 SerGlySerGlnIleTrpProGlyLeuSerLeuAspLeuLysSerGlyValValLys
234 AGCTGCATCAGTCAAACACCGGSGNTAAAATCTNGAATCGGATTGGTTCCGGGCAGGTGAAG

141 IleLeuProLysGluGlnLeuHisCysLysMetProLeuAlaSerGlySerHisProGlyVal
244 ATAATACCTAAAGAGSGAACACTGAAAATGCNGAAGCAGTTCAAAGTCAAAGGTCACAAGTT

```
AX369228
LOCUS       AX369228                 486 bp    DNA     linear   PAT 16-FEB-2002
DEFINITION  Sequence 1938 from Patent WO0204514.
ACCESSION   AX369228
VERSION     AX369228.1  GI:18857206
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Wang,T., Watanabe,Y., Henderson,R.A., Johnson,J.C., Retter,M.W.,
            Marnerakis,M., Carter,D., Fanger,G.R., Vedvick,T.S., Bangur,C.S.,
            Mcnabb,A., Fanger,N., Switzer,A., Mcneill,P.D. and Clapper,J.D.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: WO 0204514-A 1938 17-JAN-2002;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1..486
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      150 a    118 c    149 g     69 t
ORIGIN

Alignment Scores:
Pred. No.:                 9.57e-55        Length:          486
Score:                     887.00          Matches:         160
Percent Similarity:        100.00%         Conservative:    0
Best Local Similarity:     100.00%         Mismatches:      0
Query Match:               100.00%         Indels:          0
DB:                        6               Gaps:            0

US-10-017-754-809 (1-160) x AX369228 (1-486)

Qy          1 MetArgCysHisAlaHisGlyProSerCysLeuValThrAlaIleThrArgGluGluGly 20
                |||:|||| |||  |||: .||: |||:  |||' |||: |||'  :|||  |||
Db          1 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAGGGAGGAAGGA 60

Qy         21 GlyProArgSerGlyGlyAlaGlnAlaLysLeuGlyCysCysTrpGlyTyrProSerPro 40
               .||| :|||: |||  .||: .||   |||| .||': |||  .|||  |||
Db         61 GGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCTGTTGGGGGTATCCGAGTCCC 120

Qy         41 ArgSerThrTrpAsnProAspArgArgPheTrpThrProGlnThrGlyProGlyGluGly 60
               ||  .|||  |||. '||:  '|||   :|||  ||||  |||  |||  .||'  |
Db        121 AGAAGCACCTGGAACCCGGACAGAAGATTCTGGACTCCCCAGACGGGACCAGGAGAGGGA 180

Qy         61 ArgHisGluArgHisThrGlnThrGlnAsnHisThrAlaSerProArgSerProValMet 80
               |||'  ||||: .||||  .|||   |||  .|||  |||  .|||'  .|||
Db        181 CGGCATGAGCGACACACACAAACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATG 240

Qy         81 GluSerProLysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 100
               |||  .|||'  .|||'  .|||'  .|||:'  .||||:  |||'  .|||'  .||
Db        241 GAGAGCCCCAAAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 300

Qy        101 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysValIleCysLys 120
               .:||'  .|||'  .||||   |||   ||:'  .|||  ||||  |||  .||
Db        301 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGTCATCTGCAAG 360

Qy        121 SerCysIleSerGlnThrProGlyIleAsnLeuAspLeuGlySerGlyValLysValLys 140
               : '|||:. '||||. |||  |||'  ||||  |||:  ||||  ||||  |||  .||
Db        361 AGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGGGTTCCGGCGTCAAGGTGAAG 420

Qy        141 IleIleProLysGluGluHisCysLysMetProGluAlaGlyGluGluGlnProGlnVal 160
               |'  .|||  .|||'. .|||   |||   :|||  :|||   ||||.·|||·  |||
Db        421 ATAATACCTAAAGAGGAACACTGTAAAATGCCAGAAGCAGGTGAAGAGCAACCACAAGTT 480

RESULT 2
AX369231
LOCUS       AX369231                 486 bp    DNA     linear   PAT 16-FEB-2002
DEFINITION  Sequence 1941 from Patent WO0204514.
ACCESSION   AX369231
VERSION     AX369231.1  GI:18857209
```

```
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Wang,T., Watanabe,Y., Henderson,R.A., Johnson,J.C., Retter,M.W.,
            Marnerakis,M., Carter,D., Fanger,G.R., Vedvick,T.S., Bangur,C.S.,
            Mcnabb,A., Fanger,N., Switzer,A., Mcneill,P.D. and Clapper,J.D.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: WO 0204514-A 1941 17-JAN-2002;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1..486
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      150 a    118 c    149 g     69 t
ORIGIN

Alignment Scores:
Pred. No.:                 9.57e-55        Length:          486
Score:                     887.00          Matches:         160
Percent Similarity:        100.00%         Conservative:    0
Best Local Similarity:     100.00%         Mismatches:      0
Query Match:               100.00%         Indels:          0
DB:                        6               Gaps:            0

US-10-017-754-809 (1-160) x AX369231 (1-486)

Qy          1 MetArgCysHisAlaHisGlyProSerCysLeuValThrAlaIleThrArgGluGluGly 20
               .||| :.||| '||||' |||| '||||| : '||||:||||·'|||:'|||  :|
Db          1 ATGCGGTGCCACGCCCATGGACCTTCTTGTCTCGTCACGGCCATAACTAGGGAGGAAGGA 60

Qy         21 GlyProArgSerGlyGlyAlaGlnAlaLysLeuGlyCysCysTrpGlyTyrProSerPro 40
              |', |||'. .|||| ',|||:;|||||| ',||||'!'||||.;||||, |||  |||
Db         61 GGGCCGAGGAGTGGAGGGGCTCAGGCGAAGCTGGGGTGCTGTTGGGGGTATCCGAGTCCC 120

Qy         41 ArgSerThrTrpAsnProAspArgArgPheTrpThrProGlnThrGlyProGlyGluGly 60
              ;||'. |||  .;||: ',|||'.',||||',,||||:,;|||·',||| :'||' :||
Db        121 AGAAGCACCTGGAACCCGGACAGAAGATTCTGGACTCCCCAGACGGGACCAGGAGAGGGA 180

Qy         61 ArgHisGluArgHisThrGlnThrGlnAsnHisThrAlaSerProArgSerProValMet 80
              ,,|||' ',|||' ,|||| ',|||',',|||||.,.||||'; ||||.',|||' |||' ,
Db        181 CGGCATGAGCGACACACACAAACACAGAACCACACAGCCAGTCCCAGGAGCCCAGTAATG 240

Qy         81 GluSerProLysLysLysAsnGlnGlnLeuLysValGlyIleLeuHisLeuGlySerArg 100
              | ';|||:  ;||| . |||| '||||:,'||||| |'||||',:,|||';,||':;||:
Db        241 GAGAGCCCCAAAAAGAAGAACCAGCAGCTGAAAGTCGGGATCCTACACCTGGGCAGCAGA 300

Qy        101 GlnLysLysIleArgIleGlnLeuArgSerGlnCysAlaThrTrpLysValIleCysLys 120
              ||  ,|||: ;|||! ;|||!  ||||  :,||||;,!:||||| : ;||||': |||'
Db        301 CAGAAGAAGATCAGGATACAGCTGAGATCCCAGTGCGCGACATGGAAGGTCATCTGCAAG 360

Qy        121 SerCysIleSerGlnThrProGlyIleAsnLeuAspLeuGlySerGlyValLysValLys 140
              ,||! ,||' ,||' ;|||' ,||||! ::,|||||, ||||' ',|||  .||'
Db        361 AGCTGCATCAGTCAAACACCGGGGATAAATCTGGATTTGGGTTCCGGCGTCAAGGTGAAG 420

Qy        141 IleIleProLysGluGluHisCysLysMetProGluAlaGlyGluGluGlnProGlnVal 160
              .|||' ,||'  |||'.'|||''.||||''::|||||,!||||||!  .,|||' .|
Db        421 ATAATACCTAAAGAGGAACACTGTAAAATGCCAGAAGCAGGTGAAGAGCAACCACAAGTT 480

RESULT 3
HSAJ18878
LOCUS       HSA318878                740 bp    mRNA    linear   PRI 14-JAN-2002
DEFINITION  Homo sapiens mRNA for XAGE-1c protein.
ACCESSION   AJ318878
VERSION     AJ318878.1  GI:18157209
KEYWORDS    XAGE-1c gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

───────────────────────────────

REFERENCE 1
  AUTHORS   Mulder, A.J.W., de Knots, A.A., Weidle, U.H., Radis, G.J. and van
  TITLE   Expression profile of members of the MAGE cancer/testis antige...
  JOURNAL   Unpublished

REFERENCE 2 (bases 1 to 782)
  AUTHORS   Mulder, A.J.W.
  TITLE   Direct Submission
  JOURNAL   Submitted (14-AUG-2001) Boesdman A.J.W., Department of Pathology,
            University Medical Center St Radboud, Geert Grooteplein Zuid 24, P.
            Box 9101 6500 HB Nijmegen, NETHERLANDS

FEATURES   Location/Qualifiers
     source   1..782
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xq11-q13"
                     /clone="XAGE-1"
                     /tissue_type="..."
     gene   1..745
                     /gene="XAGE-1"
     CDS   ...
                     /codon_start=1
                     /product="XAGE-1"
                     /protein_id="..."
                     /translation="M..."

BASE COUNT
ORIGIN

US-10-017-754-809 (1-180) x AX063181 (1-)

Alignment Scores:
Pred. No: 1
Score: 1480
Percent Similarity: 897.00
Percent Identity: 100.00
Best Local Similarity: 100.00%
Query Match: 100.60%

Length: 740
Matches: 740
Conservation: 0
Mismatches: 0
Indels: 0
Gaps: 0

[sequence alignment rows - illegible]

─────────────────────────────────

US-10-017-754-809.rge

REFERENCE 1
  AUTHORS   ...
  TITLE   ...
  JOURNAL   ...

ID   782 ATG...

FEATURES   Location/Qualifiers
     source   ...
                     /organism="Homo sapiens"
                     /mol_type="mRNA"

BASE COUNT
ORIGIN

US-10-017-754-809 (1-180) x AX063181 (1-781)

Alignment Scores:
Pred. No: 1
Score: 887
Percent Similarity: 100.00
Percent Identity: 100.00
Best Local Similarity: 100.60%
Query Match: 100.004

Length: 782
Matches: 782
Conservation: 0
Mismatches: 0
Indels: 0
Gaps: 0

[sequence alignment rows - illegible]

DEFINITION   Sequence 809 from Patent WO254514.
ACCESSION   AX369094
VERSION   AX369094
KEYWORDS
SOURCE
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  AUTHORS   Wand,I., Watanabe,Y., Henderson,A., Johnson,J.G., Retter,M.W.,
             Matsuraka,S.M., Carter,P., Fanger,G.R., Vedvick,T.S., Hagen,S.R.,
             Mohaoto,A., Fanger,N., Saitguca,R., Maxwell,P.G. and Clapper,J.G.
             Mammalian Euthoria; Primates; Catarrhini; Hominidae; Homo.
  TITLE   Compositions and methods for the therapy and diagnosis of lung
             cancer
  JOURNAL   Patent: WO 0204214-A 809 17-JAN-2002;
             CORIXA CORPORATION (US)

FEATURES   Location/Qualifiers
     source   1..781
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"

BASE COUNT   229 a   191 c   254 g   95 t
ORIGIN

US-10-017-754-809 (1-809) x AX369094 (1-781)

Alignment Scores:
Pred. No.:   1578..54
Score: 887.95
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 5

Length: 781
Matches: 160
Conservative: 0
Mismatches: 0
Index: 0
Gaps: 0

[sequence alignment block — illegible]

ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  AUTHORS   Wand,I., Watanabe,Y., Henderson,A., Johnson,J.G., Retter,M.W.,
             Matsuraka,S.M., Carter,P., Fanger,G.R., Vedvick,T.S., Hagen,S.R.,
             Mohaoto,A., Fanger,N., Saitguca,R., Maxwell,P.G. and Clapper,J.G.
  TITLE   Compositions and methods for the therapy and diagnosis of lung
             cancer
  JOURNAL   Patent: WO 0204214-A 944 17-JAN-2002;
             CORIXA CORPORATION (US)

FEATURES   Location/Qualifiers
     source   1..507

BASE COUNT   157 a   125 c   150 g   74 t
ORIGIN

Alignment Scores:
Pred. No.:   2.26..54
Score: 982.90
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match: 98.44%

Length: 509
Matches:
Conservative: 0
Mismatches: 0
Index: 0
Gaps: 0

US-10-017-754-809 (1-809) x AX369222 (1-507)

[sequence alignment block — illegible]

RESULT 3
AX369212
DEFINITION   Sequence 1922 from Patent WO254514.
ACCESSION   AX369212
VERSION   AX369212.1   GI:18857260
KEYWORDS
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  AUTHORS   Strathorn,B.

(sequence alignment data — illegible)

RESULT 13

LOCUS	AF251237	626 bp	mRNA	linear	PRI 23-AUG-2000

DEFINITION Homo sapiens XAGE-1 mRNA, complete cds.
ACCESSION   AF251237
VERSION   AF251237.1 GI:9885424
KEYWORDS   .
SOURCE   Homo sapiens.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 626)
  AUTHORS   Liu,X.F., Helman,L.J., Yeung,C., Lee,B. and Pastan,I.
  TITLE   XAGE-1, A New Gene that is Frequently Expressed in Ewing's Sarcoma
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 626)
  AUTHORS   Liu,X.F., Helman,L.J., Yeung,C., Lee,B. and Pastan,I.
  TITLE   Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Lab of Mol. Biol., NCI, NIH, 37 Convent Dr.
            MSC 4255, Bethesda, MD 20892, USA
FEATURES   Location/Qualifiers
     source   1..626
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /tissue_type="testis"
     CDS   66..526
                     /note="expressed in Ewing's sarcoma"
                     /codon_start=1
                     /product="XAGE-1"
                     /protein_id="AAG01401.1"
                     /db_xref="GI:9885425"
                     /translation="MLIWCFPQCACSCGVFPSAFSPVWGIRRPCFHATRVPEVWILSH
                     LREGGHTCTQNHIASPRSPVHESPKKKNQQIKVILHLGSRJKKIFICLRSQRATWR
                     VICKSRSRSQFFSINLDLSSGSVKYIPKEEHCKMPEAEEEQPCV"
BASE COUNT   154 a   170 c   199 t   g
ORIGIN

US-10-017-754-809 (1-626) x AF251237 (1-626)

Alignment Scores:
Prob. No.:   5.4e-29
Score: 524.00
Percent Similarity: 95.97%
Best Local Similarity: 94.68%
Query Match: 97.08%

Length: 626
Matches: 115
Conservative: 2
Mismatches: 27
Indels: 11
Gaps: 4

```
Qy     16 ThrArgGluGluGlyGlyHisThrCysThrGlnAsnHisIleAlaSerProArgSerProValHisGluSerProLysLysLysLysCysCysTrp   95
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66 AC.AGGGAAGGAGGTCA.TGCGCATCCCATCCCTTTGGGGCCACCT-CAGTCGGGCATCGTCAC.S  124

Qy     96 Gly--TyrProSer---ProArgSerThrAlaSerProAsnProAspProAlaArgGlyArgPheTyrTrpThrProPro   159
           |||| |||| |||  | | |
Db    125 GGCGTGTTCCCATCGCCCGCCCGTTCGCCAGTGCTGAGGA  -ACATGGGCGAGCTGTGAGCG  181

Qy     54 GluThrGlyProGlyGlyGlu------------------------ ------ AlaHisGluArgHisThr   66
           || ||||
Db    182 GGGACTGGGATCGCCGAGGTCTGAATTCTTCTCCGTACTGGACACTGGGAGA  241

Qy     67 GlnThrCysThrHisThrAlaSerProArgSerProValMetGluSerProLysLysLys   86
           |||| || |||
Db    242 CAAACACAGCAAGCTACACACCCAGTGCCAGAGCCCCAGT-AATGGAGAGCAGCCCAAAGAAGAAG   301

Qy     87 AsnGlnHisLeuLysValIleLeuGluHisLeuGlySerArgGlnLysLysIleArgIle   116
           ||| |||||||| |||
Db    302 AACCAGCAGCTGAAAGTCGGGCTACCCTACACCTGGGCGAGACAGAAGAGGCAGAGATGACGGTA   361

Qy    107 GluLeuArgSerGlnArgAlaThrTrpArgValGlyCysLysSerCysSerIleSerSerGlnThr   126
           ||| |||||||||| |||
Db    362 CAGCTGAGATCCCAGCGGCCACATGGAGAGGAATGCAAGTCTCAAGACTGCAGCATCAGCTCAGCA   421

Qy    127 ProGlyIleAsnLeuAspLeuSerGlyValLysValIleProLysGluGluHisGlu  146
           | ||||||||||||| |||
Db    422 CCGAGGAACAATCTGGATCTCAGTGGGCTCAAGGTCAAGATAATACCTAAAGAGGAA   481

Qy    147 HisCysLysMetProGluAlaGluGluGluGlnProCysVal  160
           |||||||||||||||||||||||||||||||
Db    482 CACTGTAAAATGCCAGAAGCAGTGAAGAGGAACAACACAGTT  523
```